Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

A99.1
F764Un
copy 2

United States
Department of
Agriculture

Forest Service

**Intermountain
Forest and Range
Experiment Station
Ogden, UT 84401**

Research Note
INT-336

November 1983

Developing Prediction Models for Private Timber Harvest in Montana

U.S. DEPT. OF AGRICULTURE
NAT'L AGRIC. LIBRARY
FEB 20 '84

**Ervin G. Schuster and
Michael J. Niccolucci[1]**

Developing Prediction Models for Private Timber Harvest in Montana

Ervin G. Schuster and Michael J. Niccolucci[1]

ABSTRACT

The effects that regional size and various predictive variables have on predicting nonindustrial private timber harvest in Montana are presented. Multivariable models predicted harvest substantially better than single variable, price-based models. Multicounty area models were not consistently better predictors of harvest than the State-level model. Considerable losses in predictive ability relative to area-specific models were found when, for consistency purposes, a common model was used to predict harvest for each area.

KEYWORDS: timber harvest, nonindustrial private owners, modeling, multiple regression

When one thinks of timber harvest on nonindustrial private forest land (NIPF), visions of the eastern part of the country come to mind—the Northeast, Midwest, and South—certainly not the Intermountain West or Montana. These lands are dominated by public forests and NIPF harvest is of little concern. Right? Well, only partially so. While a decade and a half ago the NIPF sector produced about 2 percent of Montana's timber harvest, it now produces one-fifth of the State's harvest or about 200 million board feet annually (unpublished data, Northern Region, Forest Service, and Montana Division of Forestry). Montana's increased NIPF harvest reflects a more general realinement of timber supply sources in the Intermountain region—emphasize private, deemphasize public.

With the increasingly important role played by the NIPF base has come increasing interest in better understanding and predicting timber harvests from these lands. Concerns historically shown by State forestry organizations are now being expressed by the Forest Service, U.S. Department of Agriculture, where the need for forest- and regional-level planning compels more explicit attention to NIPF timber production.

Unlike the long history of NIPF-oriented investigations found elsewhere, the Intermountain West and Montana have few or no such traditions. Recently available timber harvest and supply analyses pertaining to these areas are not well-suited to needed evaluations (see USDA Forest Service 1982 and Adams and Haynes 1980). Because these analyses have been directed toward a national timber assessment with eight major regions, relationships for smaller geographical areas were not developed. Additionally, the supply data for the Rocky Mountain region applied to the total private sector, not distinguishing forest industry from NIPF.

The research reported here was conducted in response to a continuing need to predict timber harvest for all producing sectors of the State and sub-State levels. The research hypothesis asserted that substantial improvements could be made by estimating NIPF timber harvest separately from forest industry, by adopting smaller geographical regions, and by expanding the range of predictor variables available. While appearing similar to models of economic supply, models developed were formulated differently to predict timber harvest only.

METHODS

Montana was chosen for study because of its importance in Intermountain States timber production and readily available data. Figure 1 shows the boundaries of the four geographical regions used in this study. Based on available timber production and marketing characteristics, two small multicounty areas were identified—northwestern and southwestern Montana. Either contains a land base greater than Maryland or Vermont. These two areas also were aggregated into a western Montana region, with a land area about the size of West Virginia. The final region consisted of the total State, the fourth largest in the Nation.

Selection of the potential independent variables such as price was based on economic theory, while others such as ownership size were based on previous NIPF studies that identified factors affecting landowner timber harvest behavior. Five broad categories were identified: (a) characteristics of the NIPF land base, (b) economic conditions prevailing in the area, (c) price received for the stumpage, (d) timber harvest behavior of other stumpage producers, and (e) a miscellaneous category.

While theory and previous studies suggested these categories, neither adequately specified the actual variables to be used. Consequently, we selected several candidate variables for each category that had the potential of being useful in prediction. Further, we had a choice of expressing many of these candidate variables in several ways; for instance, dollar-measured candidate variables could be expressed in nominal or constant

[1]Research forester and economics assistant, respectively, at the Intermountain Station's Forestry Sciences Laboratory, Missoula, Mont.

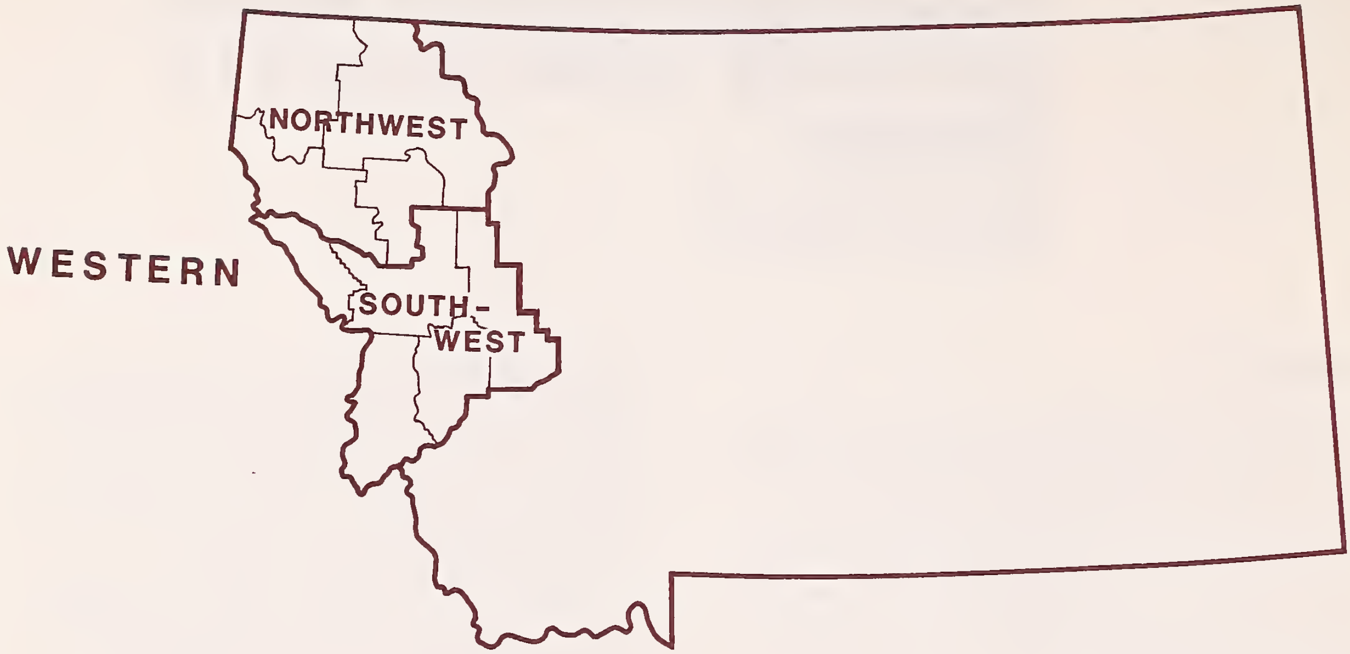

Figure 1.—Montana and three multicounty regions used to assess prediction models in this study.

(real) dollars and most variables could be expressed in current year or previous year (lagged 1 year) levels. Constant dollar conversions used the Gross National Product Implicit Price Deflator, 1972 base. We ultimately treated variables in one of three ways:

Type	Treatments
I	Nominal dollars; current year
II	Nominal and real dollars; current and previous years
III	Nominal dollars; current and previous years

Table 1 summarizes the categories of variables, types of candidate variables, and a treatment code (I, II, or III) for each. In fact, we used even more candidate variables than are indicated. For example, the per capita income variable shown was really represented by three variant forms: per capita personal income, per capita disposable personal income, and per capita nonfarm income, each treated in a Type II manner. In total, we quantified several dozen analysis variables, combinations of candidate variables, variant forms, and methods of treatment.

Data series were developed for the dependent variable, NIPF timber harvest (MMBF, Scribner), and for each of the potential independent variables, for each designated region, for the 16-year period 1965 through 1980. County-level or quasicounty-level data were generally obtained and aggregated into study regions. Savings rates, interest rates, and two price indexes are national in scope. A wide variety of data sources were used.

Data were fit to multiple linear regression models using ordinary least squares and the BMDP stepwise regression computer procedure (Dixon 1981). In linear regression, available data are used to estimate coefficients for the model's independent variables so as to account for as much variability in the dependent variable as possible. Region-specific data for all available years were used to estimate each model. Development of the "best" model proceeded until either no new variables were statistically significant ($\alpha = 0.1$) or when the three most significant variables had been identified (see Draper and Smith 1981).

For the models reported in this paper, the unrestricted, automatic stepping process of the computer

Table 1.—Categories of independent variables, candidate variables, and treatment code (I, II, or III) used to predict NIPF timber harvest

NIPF characteristics	Economic conditions	Price received	Other harvests	Miscellaneous
Acres per owner, I	Per capita income, II	$/MBF, II	National Forest, III	Lumber output, III
Volume per acre, I	Interest rate, national, III		Forest industry, III	Residual, III
Growth per acre, I	Savings rate, national, III		Uncut volume under contract, III	
Timber size, I	Unemployment rate, III			

procedure was frequently interrupted to preclude selection of "redundant" variables. Many candidate variables were handled with variant forms and methods of treatment because of uncertainty as to which form or treatment would prove most useful statistically. When one of these variables was selected by the computer program for inclusion into a model, the related variables were then irrelevant, superfluous to the analysis, and excluded from subsequent consideration. For example, if the per capita personal income variable (real dollars, current year) was selected for inclusion in a model, not only were all other methods of treatment for that variable (for example, nominal dollars, previous year) excluded from future consideration, but the alternative variant forms (such as, per capita disposable personal income and per capita nonfarm income) were excluded as well. Once the "best" models were identified, a series of "all possible combinations" of variables regression runs was made to determine if a preferable model existed but was missed by the stepwise procedure.

All "best" models were tested for statistical significance and consistency with underlying assumptions. Statistical significance of regression coefficients was determined with a two-tailed test. That procedure, which ignores the sign of the coefficient, was judged appropriate for prediction-motivated modeling. Subsequent tests for multicolinearity, heteroscedasticity, and autocorrelation either revealed no problem, or problems were resolved; one autocorrelation test was inconclusive (see Koutsoyiannis 1977). Results presented use the unadjusted coefficient of multiple determination (R^2) rather than the adjusted coefficient because of its understandability while faithfully reflecting the adjusted coefficient in these analyses.

RESULTS

Predictive equations resulting from our modeling efforts seemed quite satisfactory, accounting for three-fourths to almost 90 percent of the variation in NIPF timber harvest. Equations for these, "best" models are shown below for Montana (MT), western Montana (WM), southwestern Montana (SW), and northwestern Montana (NW) regions together with t-statistics, level of significance (α = 0.01, 0.05, or 0.10), and the percentage of variation in harvest explained by the model, the R^2.

$$H_t^{MT} = \underset{}{258.89} + \underset{(6.72,\ .01)}{4.75(P)_t} - \underset{(-2.95,\ .05)}{0.16(Lmbr)_{t-1}} \quad \ldots R^2 = 80.7$$

$$H_t^{WM} = -27.99 + \underset{(8.15,\ .01)}{2.70(P)_t} + \underset{(-2.60,\ .05)}{8.47(S)_{t-1}} \quad \ldots R^2 = 83.7$$

$$H_t^{SW} = 95.71 + \underset{(5.22,\ .01)}{0.76(P)_t} - \underset{(-5.29,\ .01)}{1.27(Res)_t} + \underset{(1.86,\ .10)}{2.4(Unemp)_t} \quad \ldots R^2 = 89.3$$

$$H_t^{NW} = -8.37 + \underset{(5.59,\ .01)}{1.30(P)_t} + \underset{(4.48,\ .01)}{0.80(S)^2_{t-1}} \quad \ldots R^2 = 76.5$$

where:

H = NIPF timber harvest; region (MMBF, Scribner).
P = National Forest stumpage price, cut; region; \$/MBF (1972 \$).
Lmbr = lumber output; region (MMBF).
S = savings rate; nation (percent = savings as percent of personal disposable income).
Res = residual; region (percent = National Forest harvest + industry harvest as percent of lumber output).
Unemp = unemployment rate; region (percent = unemployed workers as percent of labor force).
t = year.

Given the large number of candidate independent variables, the recurrence of certain variables in the "best" models is comforting, if not somewhat surprising. The real price received for National Forest stumpage cut was clearly the single best predictor variable in all four models. We do not contend that National Forest stumpage prices cause NIPF owners to cut timber; clearly they do not. Rather, because no data are available on actual prices paid to NIPF owners, any available price data can only represent an index or proxy for the actual stumpage price received by the landowners. National Forest stumpage cut prices and NIPF stumpage prices probably move together.

In addition to price, the national savings rate (S) was an important predictor variable in the western and northwestern Montana models. In both cases, the "previous year" form of the savings rate variable was used and the sign of the coefficient was positive. The residual variable (Res) was also an important predictor in southwestern and Montana models. In the "current year" form, it had a negative coefficient. Loosely interpreted, this variable represents (on a percentage basis) the proportion of the region's lumber output accounted for by timber harvested from industry and National Forest lands. The higher the proportion, the lower the NIPF harvest. It portrays NIPF harvest as filling an unmet or residual need for timber. The "previous year" form of the lumber output variable (Lmbr) was significant in the Montana model. Finally, unemployment rate was important in the southwestern Montana model, being positively related to NIPF harvest. The higher the rate of unemployment, the higher the harvest.

All "best" models accounted for a substantial portion of the variation in NIPF timber harvest, as reflected by the relatively large R^2 values. The larger this value, the more variation explained by the model, and the more useful the model is in prediction. Table 2 shows a comparison of the actual NIPF harvest and the harvest predicted by the "best" model, for each region. The "best" models vary in prediction ability. The northwestern, southwestern, and western Montana models had the smallest average, annual percentage difference between observed and predicted, each averaging just under 10 percent over the 16-year period. The Montana model had the largest differences, averaging about 13 percent per year over the period. These larger differences are, no doubt, related to the volatility of eastern Montana harvest.

Derivation of predicted harvest levels can be illustrated by the southwest model for 1980. Given the 1980 level for real price (P = \$20.87/MBF), residual (Res = 77.0 percent), and unemployment (Unemp = 7.5 percent), the level of harvest is predicted as follows:

$$H_{1980}^{SW} = 95.71 + 0.76(20.87) - 1.27(77.0) + 2.4(7.5)$$
$$= 31.4 \text{ MMBF}$$

Table 2.— Comparison of actual NIPF timber harvest and that predicted by "best" models, by region, 1965-80

Year	Northwestern Predicted	Northwestern Actual	Southwestern Predicted	Southwestern Actual	Western Predicted	Western Actual	Montana Predicted	Montana Actual
	Million board feet							
1965	43.6	59.0	19.9	18.7	64.3	77.7	112.3	124.4
1966	41.9	36.7	13.3	12.1	63.1	48.8	108.5	105.1
1967	44.7	42.8	18.5	19.1	65.1	61.8	99.8	98.4
1968	61.2	49.3	23.8	21.8	84.1	71.1	121.5	99.9
1969	70.9	63.9	35.9	45.6	117.1	109.5	160.3	151.2
1970	49.4	49.0	43.3	39.3	84.9	88.3	138.9	140.0
1971	61.7	72.5	24.6	27.1	86.5	99.6	140.3	180.1
1972	82.1	92.1	34.3	30.7	122.1	122.8	173.1	134.9
1973	69.6	75.2	56.8	53.6	118.8	128.9	200.6	215.7
1974	86.4	86.7	40.1	46.5	134.9	133.2	184.9	229.7
1975	64.6	53.3	37.8	30.7	95.2	84.0	176.2	148.7
1976	70.5	67.2	47.9	49.5	105.8	116.7	214.7	231.1
1977	61.3	61.3	56.4	64.6	110.1	125.9	222.0	218.2
1978	63.2	65.4	49.9	51.5	114.5	116.9	227.6	222.2
1979	65.9	62.9	57.7	44.9	129.9	112.9	238.3	227.9
1980	42.6	42.0	31.4	31.2	74.6	73.2	157.4	148.6

Estimates for other regions and years are similarly derived. The estimate for the 1981 harvest, a year outside the modeling data base, in the southeastern region is calculated as follows:

$$\begin{aligned} H^{SW}_{1981} &= 95.71 + 0.76(21.85) - 1.27(86.1) + 2.4(9.2) \\ &= 25.0 \text{ MMBF} \end{aligned}$$

The actual level of the 1981 harvest for that region was 25.1 MMBF. Differences between the predicted and actual harvests for the northwestern and western Montana regions were typical of the 16-year period; that for Montana exceeded the typical difference.

One purpose of our work was to see if improvements in predicting NIPF harvest could be made by expanding the range of independent variables considered. We found that although addition of the second and third variables could markedly improve prediction, choice of the first variable was critical. Table 3 shows a comparison of model explanation capability (R^2) between single variable models using several real price-based variables and the multiple-variable "best" models. As shown, improvements can be made over single, price-based variable models by including more explanatory variables. The largest improvement, a 90 percent increase, was obtained in the two-variable northwestern model, relative to the best price-based model. This point is further emphasized when the multivariable models are compared to a widely used price variable, National Forest stumpage (sold) price.

Substantial prediction improvements can be made by using the "correct" price variable. Inspection of the

Table 3.—Comparison of explanatory ability (R^2) for alternative "real" price-based models relative to "best" multivariable models, by region

Region	Alternative price measures/indexes[1]					"Best" model;[2] variables =		
	WWPA[3]	WPI[4]	Mont[5]	NFS[6]	NFC[7]	1	2	3
Montana	18.5	52.0	57.2	30.7	67.7	67.7	80.7	—
Western	0.3	22.7	49.8	55.2	75.2	75.2	83.7	—
Southwestern	7.0	44.2	49.8	50.3	55.5	55.5	86.2	89.3
Northwestern	2.7	2.5	25.4	16.8	40.2	40.2	76.5	—

[1]Equivalent models using "current" prices were substantially inferior to "real" price models.
[2]Independent variables significant at a = 0.10 level, or better.
[3]Western Wood Products Association, price index, dry Douglas-fir and larch, real.
[4]Wholesale Price Index, lumber and wood products, 1972 base.
[5]Montana Division of Forestry stumpage price, sold, \$/MBF, real.
[6]National Forest stumpage price, sold, \$/MBF, real.
[7]National Forest stumpage price, cut, \$/MBF, real.

predictive power of alternative price measures/indexes shows wide variation. The Western Wood Products Association price index is generally the worst, and the National Forest stumpage (cut) price is always the best. In all cases, real prices were found to be far more useful in prediction than their nominal price counterparts. All this suggests a need for analysts to rather carefully choose type of price variable to use.

Another purpose served in this study was to determine the extent to which improvements in predicting NIPF harvest could be secured by focusing on smaller geographical areas. Frankly, we expected substantial improvements. But this was not the case. Table 3 shows mixed results. A comparison between the "best" one-, two-, or three- variable models associated with increasingly larger regions (NW and SW, to WM, to Montana) shows no consistent pattern. Models for the smaller regions were better or worse than those for large ones, depending on the specific regions being compared and the number of variables in the models. These results are suggestive only, definitive conclusions being possible only from more extensive research involving replications. Nevertheless, we speculate that major improvements in predictability were secured by initially focusing on a single State rather than a large, multi-State region.

Timber harvest for western Montana can be predicted in one of two ways. It can be predicted directly by use of its "best" model as described earlier. It can also be predicted by summing the predicted harvests for the northwestern and southwestern regions. These summations provided a better (or at least equivalent) approximation of the actual harvest in western Montana than did the "best" model for two-thirds of the study years. Since a separate model for eastern Montana was not developed, we could not assess whether the sum of eastern plus western Montana provided a better approximation of the Montana harvest than did the Montana "best" model. We expect, however, that this would be the case.

Finally, we present results of analyses that modeled each area's timber cut by the variables contained in the "best" model for each other area. This was done to determine the extent to which the need (perceived or required) for model consistency between regions entailed a loss of predictive ability, measured by R^2. As planning efforts in forestry expand, there is a tendency to require procedures that ensure uniformity and comparability.

Results shown in table 4 indicate a mixed response again. The R^2 values shown are frequently exaggerated since they correspond to models that may contain statistically insignificant variables. The northwestern, southwestern, and western Montana models all explained an average of 72 percent of the variation in harvest while the Montana model averaged about 10 percent less. However, although the northwestern and western Montana models produced results that varied only by about 22 percent between regions, the Montana model produced results that varied by about 37 percent and the southwestern model by 43 percent. These results suggest that consistency can be achieved only at the expense of predictive power, a substantial loss in some cases. The southwestern, northwestern, and western Montana models all generated an average loss of about 14 percent per (outside) region. The Montana model has the largest loss, averaging about 23 percent per region. Regardless, either the largest or smallest decrease would generally be considered to be an appreciable loss in predictive power.

The consequence of model consistency can be shown by subsequent analysis of the southwestern model when applied to other regions. Three sets of regression models are shown, one for each region using the variables in the southwestern model:

$$H_t^{MT} = \underset{}{59.6} + \underset{(2.90,\ .05)}{4.0(P)_t} - \underset{(n.s.)}{62.7(Res)_t} + \underset{(n.s.)}{9.4(Unemp)_t} \quad \ldots R^2 = 70.5$$

$$= 99.9 + \underset{(4.64,\ .01)}{4.4(P)_t} - \underset{(n.s.)}{56.3\ (Res)_t} \quad \ldots R^2 = 68.1$$

$$= 51.9 + \underset{(5.42,\ .01)}{4.8(P)_t} \quad \ldots R^2 = 67.8$$

$$H_t^{WM} = 112.1 + \underset{(3.72,\ .01)}{1.9(P)_t} - \underset{(-1.73,\ \cong .10)}{142.4(Res)_t} + \underset{(n.s.)}{5.7(Unemp)_t} \quad \ldots R^2 = 80.2$$

$$= 78.4 + \underset{(5.98,\ .01)}{2.4(P)_t} - \underset{(n.s.)}{59.3(Res)_t} \quad \ldots R^2 = 76.9$$

$$= 33.0 + \underset{(6.52,\ .01)}{2.4(P)_t} \quad \ldots R^2 = 75.2$$

$$H_t^{NW} = 83.0 + \underset{(n.s.)}{0.8(P)_t} - \underset{(n.s.)}{56.8(Res)_t} + \underset{(n.s.)}{0.7(Unemp)_t} \quad \ldots R^2 = 46.6$$

$$= 81.2 + \underset{(2.52,\ .05)}{0.9(P)_t} - \underset{(n.s.)}{51.0(Res)_t} \quad \ldots R^2 = 46.4$$

$$= 32.3 + \underset{(3.07,\ .01)}{1.1(P)_t} \quad \ldots R^2 = 40.2$$

Table 4.—Comparison of explanatory ability (R^2) of "best" model for each region applied to other regions

Region "best" model for	"Best" model applied to				
	Northwestern	Southwestern	Western	Montana	Average
Montana	43.5	62.2	75.3	80.7	65.4
Western	76.3	61.2	83.7	67.9	72.3
Southwestern	46.6	89.3	80.2	70.5	71.7
Northwestern	76.5	61.3	83.5	67.9	72.3

All variables, t-statistics, and significance levels are as given earlier. These analyses rather clearly show that if statistical significance (at $\alpha = 0.10$) of regression coefficients is a requirement in model selection, the three-variable southwestern model is not feasible in any case, other than the southwest. In conjunction with the other variables, the unemployment rate variable is not significant [(n.s.)] in all cases. When the data are fit to the two-variable, (P and Res) model, the residual (Res) variable is again not significant in any of the three regions. Only the price variable (P) is statistically significant in all regions. Stated differently, starting from the three-variable southwestern model, the only model that could be consistently applied to all regions (with the significant coefficient requirement) is the single-variable price model. In fact, that model is the best, consistent, single-variable model in light of all possible variables.

DISCUSSION

Three general conclusions derive from this study. First, highly satisfactory models can be developed to predict NIPF timber harvest, if the appropriate variables have been quantified. In particular, results would have been much less satisfactory without use of the "cut" version of National Forest stumpage price, expressed in real dollars. Without the detailed county-level records on NIPF timber harvest, our analysis of Montana regions simply could not have taken place. This type of record is not consistently available in all States. Second, the ability to develop better predictive models for small areas compared to larger areas was not uniform. Sometimes it worked; sometimes it did not. This suggests that the assumed advantage of regional forecasting, reducing variability, does not always occur. Third, use of the same model to predict timber harvest for different regions has both advantages and disadvantages. The clear advantages are consistency and the attendant ease in understanding the model and interpreting results. But these advantages can only be secured by a loss of predictive ability. Results presented indicate that the loss can be substantial or nearly inconsequental.

Throughout this report, the coefficient of multiple determination (R^2) has been used as the basis for comparisons and for judging model superiority. While a convenient measure, R^2 alone is probably an inadequate basis for such judgments. Other measures such as adjusted R^2, standard errors, and ease of quantifying independent variables should also be considered. Moreover, although some of the R^2 values presented were substantially different from others, some were quite close. An R^2 of 80.2 may be statistically different from 83.7, but the difference could be judged trivially small in a practical application. The ultimate determination that one model is "close enough" to another is a judgment that must be based on the needs and priorities of the user.

REFERENCES

Adams, Darius M.; Haynes, Richard W. The 1980 timber assessment market model: stumpage projections and policy simulations. For. Sci. Monogr. 22; 1980. 64 p.

Dixon, W. J., ed. BMDP statistical software. Berkeley, CA: University of California Press; 1981. 725 p.

Draper, N. R.; Smith, H. Applied regression analysis. 2d ed. New York: John Wiley and Sons; 1981. 709 p.

Koutsoyiannis, A. Theory of econometrics. 2d ed. New York: MacMillan; 1977. 681 p.

U.S. Department of Agriculture, Forest Service. An analysis of the timber situation in the United States 1952-2030. Res. Rep. 23. Washington, DC: U.S. Department of Agriculture, Forest Service; 1982. 499 p.

U.S. GOVERNMENT PRINTING OFFICE: 1983—776-032/1046 REGION NO. 8